



FEMS Microbiology Letters 154 (1997) 337-345

Glycerol conversion to 1,3-propanediol by Clostridium pasteurianum: cloning and expression of the gene encoding 1,3-propanediol dehydrogenase

Frauke Luers, Markus Seyfried, Rolf Daniel, Gerhard Gottschalk *

Institut für Mikrobiologie der Georg-August-Universität, Grisebachstr. 8, D-37077 Göttingen, Germany Received 23 July 1997; revised 28 July 1997; accepted 28 July 1997

Abstract

When grown on glycerol as sole carbon and energy source, cell extracts of Closiridium pasteurianum exhibited activities of glycerol dehydrogenase, dihydroxyacetone kinase, glycerol dehydratase and 1,3-propanediol dehydrogenase. The genes encoding the latter two enzymes were cloned by colony hybridization using the dhalf gene of Citrobacter freundit as a heterologous DNA probe and expressed in Escherichia coli. The native molecular mass of 1,3-propanediol dehydrogenase (DbaT) is 440 000 Da. The dhaT gene of C. pasteurianum was subcloned and its nucleotide sequence (1158 bp) was determined. The deduced gene product (41 776 Da) revealed high similarity to DhaT of C. freundii (80.5% identity; 89.8% similarity).

Keywords: Clostridium pasteurianum; 1,3-Propanediol dehydrogenase; 1.3-Propanediol; Glycerol fermentation; Type III alcohol dehydrogenase: Glycerol dehydratase

1. Introduction

It has been known for about 60 years that glycerol is fermented by facultatively anaerobic bacteria to 1,3-propanediol, ethanol, 2,3-butanediol, acetic and lactic acids. Of these substances 1,3-propanediol is of industrial interest as a monomer for light-insensitive plastics, and some strains indeed form this diol as the main product. Suitable production organisms belong to the enterobacterial genera Klebsiella and Citrobacter [1]. Recently, it has been shown that some clostridial species also convert glycerol to 1,3-propanediol [2]. The fermentation pattern is different in that the clostridia form butyric acid as a by-product. Some strains of Clostridium pasteurianum produce considerable amounts of butanol and ethanol in addition [3].

The key enzymes and the corresponding genes for glycerol fermentation have been identified and charactorized only in Citrobacter freundii and Klebsiella pneumoniae [4,5]. In the absence of an external oxidant, glycerol is consumed by a dismutation process involving two pathways. Through one pathway glycerol is dehydrogenated by an NAD+-linked glycerol dehydrogenase (DhaD) to dihydroxyacetone. This product is then phosphorylated by dihydroxyacetone kinase (DhaK) and funnelled to the central metabolism [6]. Through the other pathway glycerol is de-

0378-1097/97/\$17.00 © 1997 Federation of European Microbiological Societies. Published by Elsevier Science B.V. PH S0378-1097(97)00351-0

^{*} Corresponding author. Tel.: +49 (551) 393781; Fax: +49 (551) 393793; E-mail: ggottsc@gwdg.de

hydrated by coenzyme B₁₂-dependent glycerol dehydratase (DhaB, DhaC, DhaE) to form 3-hydroxypropionaldehyde, which is reduced to the major fermentation product 1,3-propanediol by the NADHlinked 1,3-propanediol dehydrogenase (DhaT), thereby regenerating NAD+ [7,8]. The four key enzymes of this pathway are encoded by the dha regulon, the expression of which is induced when dihydroxyacetone or glycerol is present. Recently, we have cloned and expressed the entire dha regulon of C. freundii in Escherichia coli [5]. The genes encoding the four key enzymes and the corresponding gene products have been sequenced and purified [6-8]. In contrast to the 1,3-propanediol-forming enteric bacteria only little is known about the enzymes responsible for glycerol breakdown by clostridia. The activity of glycerol dehydrogenase, glycerol dehydratase and 1,3-propanediol dehydrogenase has been determined in crude extracts of C. butyricum [9] and the latter activity in C. pasteurianum [2]. To our knowledge, the genes encoding key enzymes involved in glycerol conversion to 1,3-propanediol by clostridia have not been identified and sequenced.

In this report, we describe the cloning and expression in *E. coli* of the genes encoding glycerol dehydratase and 1,3-propanediol dehydrogenase of *C. pasteurianum* and the sequence of the *dhaT* gene.

2. Materials and methods

2.1. Bacterial strains and vectors

C. pasteurianum DSM 525 and C. freundii DSM 30040 were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Braunschweig, Germany). E. coli ECL707 [4] and DH5\(\alpha\) [10] were used as hosts, and the cosmid pWE15 and the plasmid pBluescript SK+ (Stratagene GmbH, Heidelberg, Germany) were employed as the vectors for cloning experiments.

2.2. Media and growth conditions

C. pasteurianum was grown in a minimal medium according to Kell et al. [11] with 100 mM glycerol as carbon source and C. freundii as described previously [5]. E. coli was routinely cultivated at 30°C in LB

medium [10], which was supplemented with ampicillin (100 µg ml⁻¹) when necessary. Recombinant *E. coli* strains used for expression of the genes involved in glycerol breakdown were grown as described previously [6]. Fermentations were done in Hungate tubes or anaerobic flasks and media were gassed with N₂ for 30 min before sterilization. A modified MacConkey agar (lactose was replaced by 70 mM glycerol) was used to identify glycerol-utilizing recombinant *E. coli* strains.

2.3. Molecular procedures

Chromosomal DNA from C. pasteurianum was isolated applying the method of Marmur [12], partially digested with EcoRI or HindIII, and ligated into the above mentioned vectors. Digestion with restriction endonuclease, ligation, packaging of DNA, transduction of cosmids, transformation of plasmids and isolation of recombinant vectors were done according to standard procedures [10]. Transductants were screened on MacConkey-glycerol-ampicillin agar for glycerol utilization, which was indicated by a red color of the colonies.

The subcloning of genes involved in glycerol fermentation of *C. pasteurlanum* was performed in heterologous hybridization studies using the *dhaT* gene of *C. freundii* as a probe. As source for the isolation of this gene the recombinant cosmid pRD1 was used, which harbors the entire *dha* regulon of *C. freundii* [5]. Colony hybridization, Southern transfer of DNA fragments to nylon membranes and detection of ³²P-labelled probes were done according to Ausubel et al. [10]. DNA sequence was determined by the chain termination method of Sanger et al. [13]. The fidelity of the DNA sequence determined for the insert of pFL2 was confirmed by commercial sequencing (Seqlab, Göttingen, Germany).

2.4. Preparation of cell extracts

Cells of the stationary growth phase from 500 ml anaerobic cultures were harvested by centrifugation at $6000\times g$ for 20 min, washed once with 100 mM potassium phosphate buffer (pH 8.0) and resuspended in 2-3 ml of the same buffer. The cells were disrupted by French pressing $(1.38\times10^8 \text{ Pa})$ and the extract was cleared by centrifugation at

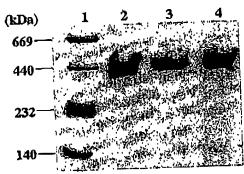


Fig. 1. Nondenaturing polyacrylamide gel electrophoresis and activity staining of 1,3-propanediol dehydrogenasa. The crude extracts were subjected to electrophoresis under nondenaturating conditions on polyacrylamide gradient slab gels (4-28%). The protein bands were stained as described in Section 2. Lanes: 1, molecular mass markers; 2, crude extract of C. freundii; 3, crude extract of C. pasteuriamum; 4, crude extract of E coli ECL707/pFL1.

 $32\,000\times g$ for 35 min at 4°C. All steps were done under anaerobic conditions.

2.5. Enzyme assays

Glycerol dehydrogenase was assayed by the method of Ruch et al. [14] and dihydroxyacetone kinase by the method of Johnson et al. [15]. Glycerol dehydratase was estimated by the 3-methyl-2-benzothiazolinone hydrazone method [16] in 1 min assays with glycerol as substrate. The activity of 1,3-propanediol dehydrogenase was determined as described previously [7]. Protein concentrations were measured by the method of Bradford [17] with bovine serum albumin as standard. All enzyme activities are expressed in µmol min⁻¹ mg protein⁻¹.

2.6. Determination of molecular mass

Electrophoresis under nondenaturing conditions was carried out on polyacrylamide gradient slab gels (4-28%) at 4°C in Tris-glycine buffer (pH 8.3) by the method of Andersson et al. [18]. Activity staining of 1,3-propanediol dehydrogenase was performed as described by Boenigk [19]. For calculation of the native molecular mass, a commercial high-molecular-mass calibration kit of standard proteins was used.

3. Results and discussion

When grown in minimal medium with 100 mM glycerol as the energy and carbon source in batch culture, C. pasteurianum formed 1,3-propanediol, butanol and ethanol as the major fermentation products (data not shown). The four key enzymes, which are known to be responsible for the conversion of glycerol to 1,3-propanediol in enteric bacteria, could be detected. The specific activities determined for glycerol dehydrogenase, dihydroxyacetone kinase, glycerol dehydratase and 1,3-propanediol dehydrogenuse in cell extracts of C. pasteurianum were 4.5, 0.1, 2.2 and 1.7 U mg⁻¹, respectively. These activities were in the same range as in cell extracts of C. freundii and E. coli ECL707/pRD1, which harbors the entire dha regulon of C. freundit (Table 1). This result indicated that C. pasteurianum ferments glycerol like the 1,3-propanediol-forming enteric bacteria by a dismutation process. This is in accordance with the pathway postulated for the glycerol fermentation of C. pasteuriaman by Dabrock et al. [3].

A genomic library of C. pasteurianum was pre-

Table 1
Specific activities of the enzymes responsible for glycerol fermentation in C. pasteuriamum*

Organism	Specific activity (µmol min ⁻¹ mg protein ⁻¹)			
3.60000.	Glyccrol dehydrogenase	Dihydroxyacetone kinase	Glycerol debydratase	
		0.10	2.2	1.7
. pasteurianum	4.5	0.09	1.1	0.9
, freundii	4.3	0.12	1.5	0.8
coli BCL707/pRD1	5.4		_b	< 0.1
coli ECL707	< 0.01	< 0.01	1.4	1.2
coli ECL707/pFL1	< 0.01	< 0.01		

^{*}Cultures were grown at 30°C and cell extracts were prepared as described in Section 2.

[&]quot;No detocrable enzyme activity.

NO. 3006

340

P. Luers et al. I FEMS Microbiology Letters 154 (1997) 337-345

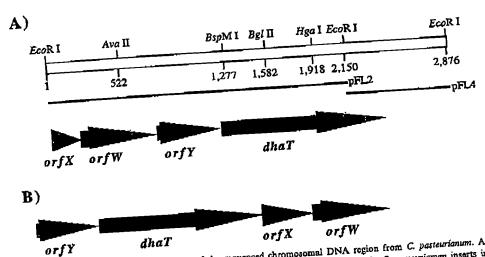


Fig. 2. A: Restriction map and genetic organization of the sequenced chromosomal DNA region from C. pasteurianum. Arrows and zerowheads represent length, location and orientation of potential genes. The location of genomic C. pasteurianum inserts in recombinant powers represent length, location and orientation of potential genes. The location of the homologous DNA region from C. freezerowheads used for sequencing is given below the restriction map. B: Genetic organization of the homologous DNA region from C. freezerowheads used for sequencing is given below the restriction map.

pared for cloning of the genes involved in glycerol breakdown. Chromosomal DNA was partially digested with EcoRI or HindIII, and ligated into the cosmid pWE15, which had been linearized with the corresponding enzymes. Ligated DNA was packed in vitro into the bacteriophage & and transduced into the glycerol minus mutant, E. coli ECL707. Approximately 2800 recombinant E. coli strains with an average insert size of 15 kb were screened on Mac-Conkey-glycerol-ampicillin agar for glycerol utilization. None of these clones had the ability to consume glycerol. This was surprising because this method had been successfully applied for cloning of the entire dha regulon from C. freundii [5]. Alternatively, the identification of the desired clones in the genomic library was done by colony hybridization using the dhaT gene of C. freundii as a heterologous DNA probe (data not shown). In this way one clone (E.

coli ECL707/pFL1) exhibiting glycerol dehydratase and 1,3-propanediol dehydrogenase activity was obtained. The recorded specific activities of 1.4 and 1.2 U mg⁻¹, respectively, were slightly lower than in C pasteurianum, but exceeded those of C. freundii (Table 1). Separation of crude extracts by gradient gel polyacrylamide electrophoresis under nondenatums conditions and activity staining of 1,3-propanediol dehydrogenase gave a single band, corresponding to a native molecular mass of 440 000 Da (Fig. 1). The 1,3-propanediol dehydrogenase produced in E coli ECL707/pFL1 was indistinguishable from the C. pasteurianum enzyme with respect to the molecular mass (Fig. 1, lanes 3 and 4). Thus, the genes encoding the reductive branch of glycerol fermentation by C. pasteurianum were cloned in E. coli ECL7071 pFL1. The recombinant cosmid recovered from this strain was designated pFL1 and contained a 13.5-kb

Fig. 3. Nucleotide sequence of the cloned region. Only one strand is shown. The gene encoding 1,3-propanediol debydrogenase (dhaT) and the putative genes or fW, or fX and or fY have been translated using the one-letter amino acid code; amino acid symbols are written below the first nucleotide of the corresponding codon. Potential ribosome binding sites (SD) and putative σ^{70} -dependent promoters are underlined. The putative secondary structure is marked by open arrows indicating the length and orientation of the stem. The sequence has been submitted in full length to GenBank under accession number AF006034.

341

F. Luers et al. I FEMS Microbiology Letters 154 (1997) 337-345
F. Lucis et al. 7 2365 155 155 155 155 155 155 155 155 155 1
1 GAATTCCAAGTTAGAAGTATAGGTATTGGAACAGATAATAATAGTCATTCAT
1 GAATTCCAAGTTAGAAGTTAGGTATAGGTATTGGAACACATTAGAACACACATTAGAACACATTAGAACACATTAGAACACACAC
OFFE
61 TTCARATTAACTTTAAATAATCCATTATTTAAACTTAAAATTAAAATTAAATTAAAATTAAAATTAAAATTAAAA
evi. The ner -
AND AND AND AND COUNTY AND
1 R F I G A N A A B L V K K N P F K N M
LEI GGATTTCATGTATTAGGTGTGAAGAAGTAGTAGAAGTGGTGATAAGG 240
ISI GGATTYCATGTATTAGGTGGTGAAGAAGTATCAGTATATAGTATATAGTATAGTATAGTAGTAGTAGTAGTA
T F M Y *
DEPARTMENT 300
241 GAGARACTGGTITATTTGGAGGAAGCAGATTAATAAAATAAAATA
A th C T k G O O T T T T
360
301 ATGGATGTTTAGATGAAGCAAATTCCTTTATAGGTCTTCTTATAGGTCTTCTTATAGGTCTTCTTATAGGTCTTCTTTATAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTAGGTCTTTTAGGTCTTTTAGGTCTTTAGGTCTTTTAGGTCTTTTAGGTCTTTAGGTCTTTAGGTCTTTTAGGTCT
ech Deans. "
A DESCRIPTION OF A CONTROL OF A
361 AAGATATAACATTATCTTAAGAAATATTCAGGACTAAAATTTTAAAAATATTAAAAAATATTAAAAAATATTAAAA
DIKTILKMIE
421 TTGCCAGTGATGAAAAGGGAAAAGCCTATCTAAAAGATACAATATCACAAGGGGATATTG 480
421 TTGCCAGTGATCAAAAGGGAAAAGCCTATCTAAAACATACAA
A S D B K G A A
491 ANGAATTGGAAAAGATTATAGATAGATATAGAGAAATTGTGGGAAAGATTGGGAAAAGATTGTGGAAAAGATTGTGGGAAAAGATTGTGGGAAAATTGGGGAAAATTGTGGGAAAATTGTGGGAAAATTGTGGGAAAATTGTGGGAAAATTGTGGGAAAATTGGGGAAAATTGTGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGGAAAATTGGGAAAATTGGGAAAATTGGGAAAATTGGGAAAATTGGGAAAATTGGGAAAAATTGGAAAATTGGAAAATTGGAAAATTGGAAAATTGGAAAAATTGGAAAAATTGGAAAAATTGGAAAAATTGGAAAAAA
E L E K I I D "
541 TTATTCCAGGTGATACAATTTCATCAGCATCATTACA
I P G D 1 1 5 5
601 GATCAGAAAGATTAATGGTGGCCTTAAAAAGCARATTAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAAG
720
661 AATATATAAATAGATTATCAGATGFFTTGTTTATACTFGCAAGAGTAGAAGCAGAAACAA 720
v 1 N R L D L v
721 ATAGAAGTTAGAAA <u>GGAAG</u> TATATCATGGCAATAAAATAAATGATTTTAAGCAGATAAG 780
721 ATAGAAGTTAGAAAGGAAGTATATCATGGCAATAGGAATAGGAATAGGAATAGGAATAGGAAGTATAGGAAGTATAGGAA
ps * SD # " "
OFFY 840
781 CTTAGAACAGTTAAGGAATGTGTAAGGCTGCAGAAATATATAT
T. E T V K E W
PROPERTY AND THE PROPERTY OF T
841 TTCAATAGTTTTTCAGCGGTGGATGCTGGCGGATATTAGT M L L T R M E S I V F S A V D A G G N L M L L T R M E
givfsavba
901 AAATGCATTTATAAGCAGTATAGATATAGCTGCCAATAAAGCTTTTACTGCATTAGCTTT 960
901 AAATGCATTTATAAGCAGTATAGATATAGCTAGC
N A F I S S A A A A A A A A A A A A A A A A A
1020
961 AAAACAAGGAACTCATGAAGTAACTCCAGTAATACAACCAGCAACACCAGCAACTCATGAACTCCAGTAATACAACCAGCAACACCAGCAACTCAGCAACTCCAGTAATACAACCAGCAACACCAGCAACTCAGCAACTCCAGTAATACAACCAGCAACACCAGCAACTCCAGTAATACAACCAGCAACACCAGCAACTCCAGTAATACAACCAGCAGCAACTCCAGTAATACAACCAGCAGCAACTCCAGTAATACAACCAGCAGCAACTCCAGTAATACAACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACTCCAGTAATACAACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
7 A G T B L V L C T
1021 ACAATTGACAAATAATTOTAGAAFTTCAACCTTTGGAGGAGCATTACCTATAATAGTTGA 1080
Q L T N N C R I S T F G G G L P I I V D
QLTNNCXI
1081 TOATCAAGTAGTGGCGATTGGAGTAAGTGGGGGAACTGTAGAAGAAGATATGTCTAT 1140
1081 TGATCAAGTAGTAGGTGCATTGGAGTAAGTGGGGGAAC TG AAGTAGTAGGTGGGGAAC TG AAGTAGGTGGGGAAC TG AAGTAGGTGGGGGAAC TG AAGTAGGTGGGGGGAAC TG AAGTAGGTGGAAC TG AAGTAGGTGGGGGGAAC TG AAGTAGGTGGGGGGAAC TG AAGTAGGAGGGGGAAC TG AAGTAGGAGGGGGGAAC TG AAGTAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGA
1141 TGCTAAATATGCATTAGATTCAATAAATGATGTTTAA <u>TTGTAA</u> TCATCATATTAAATAAA 1200
A K Y A L D S I N D V * -35
AKYALD
1201 TATAATITTAATITCTAGGAGGAATTATAAAATGAGAATGTATGATTITTTAGGACGAAA 1260
1201 TATAATTTAATTTCTAGGAGGAATTATAAAATCAGAA
1340
1261 TGTAAACTTTATGGGAGCAGGTGCAATAAATTAGGGGGGC R C K I L G

(Continued overleaf).

342

F. Lucra et al. | FEMS Microbiology Letters 154 (1997) 337-345

P. Latin by will
1321 COGARAGRARGCTTTARTACTTRACAGATARATTCTTGAGARATATGGAAGATGGAGCTGT 1380
1321 COGARAGARACTTTARTACTTACAGATARATTCTTGAGAAAAATATGAGAAAAAAAAAA
OKKALIVIDK PLKA
The state of the s
1381 AGCTCAAACGGTTAAATATTTAAAGAAGCAGGAATAGATGTTGGTTTY D D V A Q T V K Y I K E A G I D V A F Y D D V
1381 AGCTCAAACGT W Y I K E A G I D V A F
A Q T V K 1 1441 AGAGCCTAATCCTAAGGATACTAATGTTAGGGATTAAAAGTATATAGAAAAGAAA 1500 1441 AGAGCCTAATCCTAAGGATACTAATGTTAGGGATTAAAAGTATATAGAAAAGAAA 1500
1441 ACAGCCTAATCCTAAGGATACTAATCTTAGACATTAGAT
EPNPRDIM
TO THE PARTY OF TH
1501 CTGTGATTAATAGTTACGAGGAGGAAGTTCTCATGACTGTGGACTGTGGAGAGTACTGTGGAGAGAGA
C D L I V 1561 TATTGCAGCTACACACGCAGGCAGATCTTTATGACTATGCTGGTATAGAAACCCTTACTAA 1620 1561 TATTGCAGCTACACACGCAGGCAGATCTTTATGACTATGCTGGTATAGAAACCCTTACTAA 1620
TO THE TENNER OF THE TAXABLE PROPERTY OF TAXABLE PROPE
1561 TATTGCAGCTACACACGAGGCAGATCTTTATGACTATGC TGGTATATGCAGTACACGACGCAGAGCGAGATCTTTATGACTATGCAGGATCTTTATGACTATGACTATGCAGGATCTTTATGACTATGACTATGACTATGACTATGACTATGCAGGATCTTTATGACTATGA
1 A A T R 1621 TCCATTGCCTCCGATAGTAGCTGTAAATACAACAGCTGGAACAGGCAGG
1621 TCCATTGCCTCCQATAGTAGCTGTAAATATT T A G T G S E V T R
1621 TCCATTGCCTCCQATAGTAGCTGTAAATACAACAGCTGGAAAGGCTGGAAAAGGCTGGAAAGGCTGGAAAGGCTGGAAAGGCTGGAAAGGCTGGAAAGGCTGGAAAAGGCTGGAAAGGCTGGAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGGCTGGAAAAGCAAGGAAAGAAA
TO THE PROPERTY AND THE
TO THE TOTAL CARACACAAAACAAAACAAAGATTAAATTIGI AAATTIGI AATTIGI AATTIG
H C V I I
H C V I I I I I I I I I I I I I I I I I I
1741 TCTGCCGTTAGTATCCATTAATGATCCAATACTTATGATTAATGAT
TO THE PROPERTY OF THE PROPERT
1801 AGCAGCTACAGOAATGGATGCCTTAACTCATGCCATAGAGTCCTATATTTATT
TO THE PROPERTY OF THE PROPERT
A A T 0 1861 AAACCCAGTAACAGATGCCTTAGCTATACAAGCAATAAATTAATAGCTAACAATCTGCG 1920 N P V T D A L A I Q A I K L I A N N L R
N F 4 1 B 1 1 2 90
1921 TCAGGCAGFAGCCCTTGGAGAAAATCTGGAGAAAATATGGCCTAGGCATCACT 1980
1921 TCAGGLAGIAGOCCA R R R R R R R R R R R R R R R R R R
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z D D D D D D D D D D D D D D D D D D
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
Q A V A L G Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z Z
2040 A V A L G V L L G V V H A M A H Q L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A M L L P H V 2101 AGAACGCTATAATCTATATCAAATCCTAAGAAATTTGCAGAATTCATGGAGAATTCATGGG 2160
2040 A V A L G V L G V V H A M A H Q L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G
2040 A V A L G V L G V V H A M A H Q L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G
2040 A L G B B B B B B B B B B B B B B B B B B
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAGCCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAGCCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAGCCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F
2040 A V A L G L L L L L L L L L L L L L L L L L
1981 TCTGGCAGGAATGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGCGAGTTAATGAAGG 2280 R L S K D V G I F A S L K E M G V N E G
1981 TCTGGCAGGAATGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGCGAGTTAATGAAGG 2280 R L S K D V G I F A S L K E M G V N E G
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K P M G V N E G 2281 AGATTTGAAAGATATATGGCAAAAATGGCATTGAAAGAATGCATTCAGTAATCCAAG 2340
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K P M G V N E G 2281 AGATTTGAAAGATATATGGCAAAAATGGCATTGAAAGAATGCATTCAGTAATCCAAG 2340
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGGGAGTTAATGAAGG 2280 R L S K D V G I F A S L K R M G V N E G 2281 AGATTTGAATATATGGCAAAAATGGCATTGAAAGATGCATTCAAGTAATCCAAG 2340 D F E Y M A H M A L K D G N A F S N P R
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGGATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCAATTCAATGCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGGTAATGAAAAAGATTTAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGGATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCAATTCAATGCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGGTAATGAAAAAGATTTAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGGATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCAATTCAATGCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGGTAATGAAAAAGATTTAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTGAAATATAGGCAAAAATGGCATTGAAAGAATGCATTCAATCAA
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGTTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGACAGTATAAATGGGAGTTAATGAAGG 2280 E L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATGTTAGGCAAAATGGGATTGAAAGAATGGGAGTTAATCCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGGTAATGAAAAGGATATTAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R B A F * >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATTGAAAGGATTTCAGTAATGGAAGCAGCAGAAAAGCCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTGGGATACCAGCAAGTCTTAAAGAAATGGGACTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGAATGCAATTCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAAGATATAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGGAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGGGAGTTAATGAAGG 2280 R L S K D V G I F A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCATTCAGTAATCCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAAAGAATATAGTTAAAATATTTTAGGAAAGCATTTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATTGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGGGAGTTAATGAAGG 2280 R L S K D V G I P A S L K P M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGAATGCATTCAGTAATCCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGGTAATGAAAAAGATTAAATTATATTTTAAATGTAATCTTT 2400 K G N E K D I V K I F R P A F * 2401 CATATATATAAGTATTCCTTTGAAAATATTATCCTCGGATAACTATTAAAGGATATA 2460 2461 TTTATTAAACATTTATAAAAACATTTAAATATTTCAATGGCACTATTTGAGGATGCTTTAAGTGA 2520 2521 AATTTTTAAAGGGTTACATTTCAATACTGCACTAGCACCTATTTGAGGAAGATCTTTAATTAA
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGTTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAQAAATGGGACTTAATCAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCAATTCAAGTAATCCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAGGATATAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGTTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAQAAATGGGACTTAATCAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCAATTCAAGTAATCCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAGGATATAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATGAAAGGATTTCAGTAATGGAAGCAGCAGAAAAGGCTAATGAAGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATCTTGGGATACCAGCAAGTCTTAAAGAAATGGGACTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTGAATATATGGCAAAAATGGCATTGAAAGAATGGAAGTCATCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAAGATTTAAAATATTTTAGAAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F S 2401 CATATAATAAGTATTCCTTTGAAAATATTATCCTCGATAAGTTTATCAAGGATATTA 2460 2461 TTTATTAAACATTTATAAAAACATTAAATATTATCTCTCGATAAGGAACAATTTAATAAGTGA 2520 2581 GTATTATTGATAAAAATTATTTAGTTGAAAATTATTATGTTTAAATAAT
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGGTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H G V A N A N L L F H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N F K K F A D I A E F M G 2161 AGAGAATATGAAAGGATTTCAGTAATGGAAGCAGCAGAAAAGGCTAATGAAGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATCTTGGGATACCAGCAAGTCTTAAAGAAATGGGACTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTGAATATATGGCAAAAATGGCATTGAAAGAATGGAAGTCATCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAAGATTTAAAATATTTTAGAAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F S 2401 CATATAATAAGTATTCCTTTGAAAATATTATCCTCGATAAGTTTATCAAGGATATTA 2460 2461 TTTATTAAACATTTATAAAAACATTAAATATTATCTCTCGATAAGGAACAATTTAATAAGTGA 2520 2581 GTATTATTGATAAAAATTATTTAGTTGAAAATTATTATGTTTAAATAAT
1981 TCTGGCAGGAATGGCATTAATAATGCAAATTTAGGATATGTACATGCTATTGCCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ANTAGGAGGCCTGTATGATATGGCACATGGTGTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L F H V 2101 AGAACGCTATAATCTTATACAAATCCTAAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K F A D I A E F M G 2161 AGAACACTATAATCTTATCAGAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A H F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAGAAATGCGAGTTAATGAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGACTTTGAATATATGGCAAAAATGGCATTGAAAGAATGCAATTCAAGG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAAGAATATAGTTAAAATATTTTAGAGAAGACAATTTTAATTGTAATCTTT 2400 K G N E K D I V K I F R B A F * 2401 CATATAATAAGTATTCCTTAGAAATATATATATATCCTCGGATAATCAAGGATATA 2460 2461 TTTATTAAAGATTTATAAAAACATTAAATATATATCCTCGGATAATCATTAAGGAACCCCCTATTTAAGAACAATTTAAAGTATATATA
1981 TCTGGCAGGAATGGCATTTAATAATGCAAATTTAGGATATGTACATGCTATGGCACATCA 2040 L A G M A F N N A N L G Y V H A M A H Q 2041 ATTAGGAGGCCTGTATGATATGGCACATGTTGTTGCCAATGCCATGTTATTACCACATGT 2100 L G G L Y D M A H Q V A N A N L L P H V 2101 AGAACGCTATAATCTTATATCAAATCCTAAGAAATTTGCAGAATTCATGGG 2160 E R Y N L I S N P K K P A D I A E F M G 2161 AGAGAATATGAAGGACTTTCAGTAATGGAAGCAGCAGAAAAAGCTATAGATGCTATGTT 2220 E N I E G L S V M E A A E K A I D A M F 2221 TAGACTTTCAAAGGATGTTGGGATACCAGCAAGTCTTAAAQAAATGGGACTTAATCAAGG 2280 R L S K D V G I P A S L K E M G V N E G 2281 AGATTTTGAATATATGGCAAAAATGGCATTGAAAGATGCAATTCAAGTAATCCAAG 2340 D F E Y M A K M A L K D G N A F S N P R 2341 AAAAGCTAATGAAAAGGATATAGTTAAAATATTTTAGAGAAGCATTTAATTGTAATCTTT 2400 K G N E K D I V K I F R E A F >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

Fig. 3 (Continued).

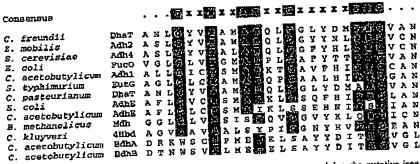


Fig. 4. Amino acid alignment of the protein regions of different alcohol dehydrogenases containing the putative iron-binding motif proposed as a typical feature of class III alcohol dehydrogenases. Shaded amino acids indicate the core motif postulated by Bairoch [21], proposed is a typical feature of class III alcohol dehydrogenases. Shaded amino acids indicate the core motif postulated by Bairoch [21], neconsensus of the putative iron-binding motif was compared with 1,3-propanediol dehydrogenase (DhaT) of C. freundii [7], alcohol dehydrogenase (Adh2) of S. cerevisiae [26], 1,2-propanediol dehydrogenase (FucO) of E coli [23], alcohol dehydrogenase (Adh1) of C. acetobutylicum [29], ethanolamine utilization protein (EutG) of S. typhomurium [25], 1,3-propanediol dehydrogenase of C. pasteurianum (DhaT), 4-bydroxybutyrate dehydrogenase (Ailbd) of C. kluyveri [30], alcohol dehydrogenase (AdhE) of E. coli and C. acetobutylicum [31,32], methanol dehydrogenase (Mdh) of Bacillus methanolicus C1 [33], and the two butanol dehydrogenase isoenzymes (BdhB and BdhA) of C. acetobutylicum [34].

insert of C. pasteurianum genomic DNA. To subclone the dhaT gene encoding 1,3-propanediol dehydrogenase, pFL1 was digested with EcoRI and the fragments were ligated into pBluescript SK⁺. Colony hybridization with the DNA probe from C. freundii (see above) revealed that the complete dhaT gene of C. pasteurianum was located on two recombinant E coli strains with different inserts, one containing a 2155-bp and the other a 732-bp EcoRI fragment of genomic C. pasteurianum DNA. The plasmids isolated from these strains were designated pFL2 and pFL4, respectively. The origin and the neighborhood on the chromosome of both cloned EcoRI fragments was established by Southern blot analysis (data not shown).

The inserts of pFL2 and pFL4 were sequenced in both directions. The restriction map and the apparent gene organization are shown in Fig. 2A, and the sequence of the combined *EcoRI* fragments from pFL2 and pFL4 (2881 bp) is given in Fig. 3. Four successive potential genes were identified within the sequence. One gene is located at the end of the cloned DNA and is hence incomplete. All presumptive genes except the incomplete one were preceded by a potential ribosome-binding site, appropriately spaced from the start codon (Fig. 3). The deduced amino acid sequences of the four open reading frames showed high similarity to OrfW, OrfX,

OrfY and DhaT, which are part of the dha regulon of C. freundii [6-8]. The C. pasteurianum genes were designated accordingly.

The dhaT gene (1158 bp) of C. pasteurianum encodes 385 amino acids with a predicted molecular mass of 41776 Da. The dhaT gene is terminated by a single stop codon (UAA). A sequence that could represent a transcriptional terminator (a punctated palindrome that could form a stem-loop structure in an RNA transcript) follows approximately 36 nucleotides downstream from the stop codon (Fig. 3). A conserved sequence for σ^{70} -dependent promoters is located upstream of the dhaT gene in positions 1178-1206 (Fig. 3).

The amino acid sequence deduced from dhaT of C. pasteurianum was compared with deduced amino acid sequences from alcohol dehydrogenases available in the NCBI databases. The highest similarity (80.5% identity and 89.8% similarity) was obtained to the 1,3-propanediol dehydrogenase of C. freundii, which is a member of a novel family of alcohol dehydrogenases (type III). This high amino acid sequence identity corresponded well with the similar native molecular mass of both enzymes observed during nondenaturing electrophoresis (Fig. 1). The 1,3-propanediol dehydrogenase of C. freundii is a decamer of a polypeptide of 43 400 Da under these conditions [19]. The predicted molecular mass of the

dhaT gene product (41776 Da) and the estimated native molecular mass (440000 Da) suggest the same subunit composition for 1,3-propanediol dehydrogenase of C. pasteurianum.

The family of type III alcohol dehydrogenases is very heterogeneous and distinct from the long-chain zinc-containing (type I) or short-chain zinc-lacking (type II) enzymes [20]. The other members of type III alcohol dehydrogenases, including e.g. Adh2 of Zymomonas mobilis and FucO of E. coli (for other enzymes, see Fig. 4), exhibited 28.3-51.6% identity (51.1-70.8% similarity) to 1,3-propanediol dehydrogenase from C. pasteurianum. No significant similarities between 1,3-propanediol dehydrogenase and type I and type II alcohol dehydrogenases were found.

Bairoch [21] proposed a more or less conserved putative iron-binding motif (G-X-X-H-X-A-H-X-X-G-X-X-X-X-X-P-H-G) as a fingerprint pattern for type III alcohol dehydrogenases (Fig. 4). It is fully conserved in all reported iron-dependent enzymes (DhaT from C. freundii [7], Adh2 from Z. mobilis [22], FucO and AdhE from E. coli [23,24]), in EutG from Salmonella typhimurium with unknown ion requirement [25] and in Adh4 from Saccharomyces cerevisiae, which requires Zn2+ for its catalytic activity [26]. The dhaT gene product showed the iron-binding motif (amino acids 262-280), except that the conserved proline in position 278 was replaced by alanine (Fig. 4). The ion requirement of the enzyme has not been determined but iron limitation during growth on glycerol favors the formation of 1,3-propanediol and reduces the production of the other solvents butanol and ethanol [3]. This makes an iron-dependent 1,3-propanediol dehydrogenase unlikely.

1,3-Propanediol dehydrogenase requires NAD(H) as a cofactor, but the highly conserved NAD(H) binding fingerprint pattern G-X-G-X-X-G [27] was not present in the amino acid sequence. This is also characteristic of most type III alcohol dehydrogenases.

The deduced products of the remaining three presumptive genes, orf Y, orf W and the incomplete orf X, exhibited 31.5-53.8% identity (45.9-69.9% similarity) to the corresponding homologous gene products encoded by the dha regulon of C. freundii. In comparison to this organism the sequenced genes of C. pas-

teurianum showed a different organization; orfX, orfW, orfY were all located upstream of the dhaT gene (Fig. 2).

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft within the Forschungsschwerpunkt 'Neuartige Reaktionen und Katalysemechanismen bei anaeroben Mikroorganismen'.

References

- [1] Homann, T., Tag, C., Biebl, H., Deckwer, W.-D. and Schink, B. (1990) Fermentation of glycerol to 1,3-propanediol by *Klebstella* and *Citrobacter* strains. Appl. Microbiol. Biotechnol. 33, 121-126.
- [2] Heyndrickx, M., De Vos, P., Vacanneyt, M. and De Ley, J. (1991) The fermentation of glycerol by Clostridium butyricum LMG 1212 t2 and 1213 t1 and Clostridium pasteurianum LMG 3285. Appl. Microbiol. Biotechnol. 34, 637-642.
- [3] Dabrock, B., Bahl, H. and Gottschalk, G. (1992) Farameters effecting solvent production by Clostridium passeurianum. Appl. Environ. Microbiol. 58, 1233-1239.
- [4] Sprenger, G.A., Hammer, B.A., Johnson, E.A. and Lin, B.C.C. (1929) Anaerobic growth of Escherichia coli on glycerol by importing genes of the dha regulon from Klebsiella pneumoniac. J. Gen. Microbiol. 135, 1255-1262.
- [5] Daniel, R. and Gottschalk, G. (1992) Growth temperature-dependent activity of glycerol dehydratase in *Escherichia coli* expressing the *Curobacter freundii dha* regulon. FEMS Microbiol. Lett. 100, 281-286.
- [6] Daniel, R., Stuertz, K. and Gottschalk, G. (1995) Biochemical and molecular characterization of the oxidative branch of glycerol utilization by Curobacter freundil. J. Bacteriol. 177, 4392-4401.
- [7] Daniel, R., Boenigk, R. and Gottschalk, G. (1995) Purification of 1,3-propanediol dehydrogenase from Citrobacter freundit: cloning sequencing and overexpression of the corresponding gene in Escherichia coli. J. Bacteriol. 177, 2151-2156.
- [8] Seyfried, M., Daniel, R. and Gottschalk, G. (1996) Cloning, sequencing and overexpression of the genes encoding comzyme B₁₂-dependent glycerol dehydratase of Citrobacter fresandii, J. Bacteriol. 178, 5793-5796.
- [9] Abbad-Andaloussi, S., Dürr, C., Raval, G. and Petitdemange, H. (1996) Carbon and electron flow in Classification butyricum grown in chemostar culture on glycerol and on glucose. Microbiology 142, 1149-1158.
- [10] Ausubel, F.M., Brent. R., Kingston, R.E., Moore, D.D., Scidman, J.O., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology. John Wiley and Sons, New York.
- [11] Kell, D.B., Peck, M.W., Rodger, G. and Morris, J.G. (1981)

- On the permeability to weak acids and bases of the cytoplasmic membrane of Clastridium partentianum. Biochem. Biophys. Res. Commun. 99, 81–88.
- [12] Marmur, J. (1961) A procedure for the isolation of desoxyribonucleic acid from microorganisms. J. Mol. Biol. 3, 208-218.
- [13] Sanger, F., Nicklen, S. and Coulson, A. R (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- [14] Ruch, F.B., Lengeler, J. and Lin, E.C.C. (1974) Regulation of glycerol catabolism in Klebsiella aerogenes. J. Bacteriol. 119, 50-56.
- [15] Johnson, E.A., Burke, S.K., Forage, R.G. and Lin, B.C.C. (1984) Purification and properties of dihydroxyacetone kinase from Klebsiella pneumoniae. J. Bacteriol. 160, 55-60.
- [16] Toraya, T., Kazutoshi, U., Fukui, S. and Hogenkamp, H.P.C. (1977) Studies on the mechanism of the adenosyl-cobalamindependent dioldchydratase reaction by the use of analogs of the coenzyme. J. Biol. Chem. 252, 963-970.
- [17] Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254.
- [18] Andersson, L.O., Borg, H. and Mikaclason, M. (1972) Molecular weight estimation of proteins by electrophoresis in polyacrylamide gel of graded porosity. FEBS Lett. 20, 199-202.
- [19] Boenigk, R. (1991) Optimierung der 1,3-Propandiolbildung mit Citrobacter fremidi und Isolierung und Charakterisierung der 1,3-Propandiol-Dehydrogenase. Doctoral Thesis, Georg-August-Universität, Göttingen.
- [20] Rsid, M.F. and Fewson, C.A. (1994) Molecular characterization of microbial alcohol dehydrogenases. Crit. Rev. Microbiol. 20, 13-56.
- [21] Bairoch, A. (1991) PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. 19 (Suppl.), 2241-2245.
- [22] Tse, P., Scopes, R.K., Wodd, A.G., Bakshi, E. and Murray, K.S. (1988) An iron-activated alcohol dehydrogenase: metal dissociation constants and magnetic and spectroscopic properties. J. Am. Chem. Soc. 110, 1295-1297.
- [23] Conway, T. and Ingram, L.O. (1989) Similarity of Escherichia coli propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from Zymomonas mobilis and Saccharomyces cerevisiae. J. Bacteriol. 171, 3754-3759.
- [24] Kessler, D., Herth, W. and Knappe, J. (1992) Ultrastructure

- and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of Escherichia coli. J. Biol. Chem. 267, 18073-18079.
- [25] Stojiljkovic, I., Bäumler, A.J. and Heffron, F. (1995) Ethanolamine utilization in Salmonella typhomatian: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB cutE cutJ cutH gene cluster. J. Bacteriol. 177, 1357-1366.
- [26] Wiliamson, V.M. and Paquin, C.E. (1987) Homology of Saccharomyces cerevisiae ADH4 to an iron-activated alcohol dehydrogenase from Zymomonas mobilis. Mol. Gen. Genet. 209, 374-381.
- [27] Wicrenga, R.K., DeMaeyer, M.C.H. and Hol, W.G.J. (1985) Interaction of pyrophosphate moieties with α-helixes in dinucleotide binding proteins. Biochemistry 24, 1346-1357.
- [28] Conway, T., Sewell, G.W., Osman, Y.A. and Ingram, L.O. (1987) Cloning and sequencing of the alcohol dehydrogenase II gene from. J. Bacteriol. 169, 2591-2597.
- [29] Youngleson, J.S., Jones, W.A., Jones, D.T. and Woods, D.R. (1988) Molecular analysis and nucleotide sequence of the adhl gene encoding an NADPH-dependent butanol dehydrogenase gene in the Gram-positive anacrobe Clastridium acetobutylicum. Gene 78, 355-364.
- [30] Söhling, B. and Gottschalk, G. (1996) Molecular Analysis of the anacrobic succinate degradation pathway in Clostridium kluyvert. J. Bacteriol. 178, 871-880.
- [31] Goodlove, P.E., Cunningham, P.R., Parker, J. and Clark, D.P. (1989) Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding gene of Escherichia coli. Gene 35, 209-214.
- [32] Fischer, R.J., Helms, J. and Dürre, P. (1993) Cloning, sequencing and molecular analysis of the rol operon of Closurdium acetobusylicum, a chromosomal locus involved in solventogenesis. J. Bacteriol. 175, 6659-6669.
- [33] De Vries, G.B., Arfman, N., Terpstra, P. and Dijkhuizen, L. (1992) Cloning, expression, and sequence analysis of the Baeillus methanolicus C1 methanol dehydrogenase gene. J. Bacteriol. 174, 5346-5353
- [34] Walter, K.A., Bennett, G.N. and Papoutsakis, E.T. (1992) Molecular characterization of two Clostridium acetobutyliaum ATCC824 butanol dehydrogenase isoenzyme genes. J. Bacteriol. 174, 7149-7158.